



GEORGE FELDMAN McDONALD, PLLC
745 Fifth Avenue, Suite 500
New York, New York 10151
Toll Free: (888) 421-4LAW
(888) 421-4529
Fax: (888) 421-4173

December 10, 2025

VIA ECF

Hon. Susan van Keulen
United State District Court for the Northern District of California
San Jose Courthouse, Courtroom 6 – 4th Floor
280 South 1st Street, San Jose, CA 95113

Re: *Smith et al. v. Google, LLC*, Case No. 5:23-cv-03527-PCP

Dear Judge van Keulen,

Pursuant to your Honor’s November 17, 2025 and November 26, 2025 orders (ECF Nos. 126, 128), I write on behalf of the parties in the above-captioned action to submit the attached table detailing the parties’ disputed and agreed upon parameters of a sampling protocol for Google’s production of sample “Class Event Data.” The parties continue to work to resolve their outstanding disputes and will promptly notify the Court if a complete resolution is reached prior to the December 16, 2025 hearing.

Respectfully yours,

/s/ Michael Liskow
Michael Liskow

Counsel for Plaintiffs

cc: Counsel of Record (via email)

Table Detailing Disputed and Agreed Upon Parameters of Sampling Protocol

Pursuant to the Court’s November 17, 2025 Order (ECF No. 126), the parties hereby identify the disputed parameters of the sampling protocol ordered by the Court, as well as fully and partially agreed upon parameters. For ease of reference, instead of identifying the fully and partially agreed parameters in this introductory paragraph, the parties have identified them in the table below.

Disputed Parameter	Plaintiffs’ Position	Google’s Position	Court’s Use
1. <i>Types of Data to be Produced</i> [AGREED]	The parties have reached agreement (memorialized in communications between the parties) on the types of data to be considered “Class Event Data” and to be produced as part of the Class Event Data sample.	[AGREED]	
1.a Sources of Data to Be Produced [TENTATIVELY AGREED]	<p>The parties have reached agreement on the specific Tax Preparer digital properties (“Properties”) for which data from the Properties will be included in the Class Event Data samples.</p> <p>Google otherwise proposes to only produce Class Event Data from two “Canonical Databases,” which respectively contain</p>	[TENTATIVELY AGREED]	

	<p>data from “Universal Analytics” and “Google Analytics 4.”</p> <p>Plaintiffs asked Google whether the Class Event Data in the Canonical Databases is entirely duplicative of the Class Event Data in any other databases, including an “upstream” database, or if there may be more (or differing) Class Event Data in other databases.</p> <p>Google’s counsel has informed Plaintiffs that, based on their investigation, (a) the two Canonical Databases contain all Class Event Data “pertinent” to end users of the tax preparation websites at issue (the end users being visitors to the Tax Preparers’ websites, including the putative class members); and (b) the “upstream” database does not contain any additional Class Event Data that pertains to the end users.</p>		
--	--	--	--

	<p>Plaintiffs accept the results of Google’s counsel’s investigation, and agree at this time to limit the source of the Class Event Data samples to the Canonical Databases (with the exception of any conversion and attribution data not contained in the Canonical Databases, as discussed above). However, Plaintiffs reserves their right to seek the purportedly non-pertinent Class Event Data from the “upstream” database if it is later determined to be relevant.</p>		
<p><i>2. Relevant Time Frame</i> [AGREED]</p>	<p>Google has represented to Plaintiffs that Class Event Data is only available for the following time periods from the agreed-to Properties:</p> <p>TaxAct: For most agreed-to TaxAct Properties, Google has data dating back to October 2023, with one Property containing data dating back to</p>	[AGREED]	

	<p>November 2017 from an earlier Google Analytics version.</p> <p>TaxSlayer: For one agreed-to TaxSlayer Property, Google has data dating back to September 2021. Two TaxSlayer Properties with data from the earlier Google Analytics version contain data dating back to 2014 and 2011 respectively.</p> <p>H&R Block: H&R Block has one agreed-to Property with data that dates back to October 2023.</p> <p>Plaintiffs accept Google's representations, without waiving their right to challenge the representations if evidence of additional time periods/ relevant properties comes to light.</p> <p>The parties therefore agree, for purposes of the sampling protocol, that the Class Event Data samples</p>		
--	---	--	--

	will be derived from the above time periods.		
3. <i>Size of Samples</i> [PARTIALLY AGREED]	<p>Google has provided Plaintiffs with the number of records, by month, for one of the two Canonical Databases. Google was unable to provide a date certain by when it would be able to produce the same information for the second Canonical Database, but Google's counsel has stated they currently anticipate that the two databases will have similar monthly metrics for each tax preparer.</p> <p>These representations are sufficient for Plaintiffs to provide a proposed sample size and sampling methodology, the elements of which are largely agreed to by the parties:</p> <p>1. Google will sample available Records for the Properties from both Canonical Databases from the time period January 1,</p>	<p>Google proposes sampling up to 5,000 records for one day each month for each relevant web property.</p> <p>The time for production depends in large part on the total number of records sampled. Accordingly, sampling a set number enables greater certainty regarding a production timeline than does a percentage, which yields a variable number of records.</p> <p>Plaintiffs request that Google provide "the total number of Records for that 24-hour day, broken down by (a) Canonical Database and (b) Relevant Property. Google will also provide the total number of Records for the entire Relevant Period, broken down by (a) Canonical Database, (b) Relevant Property and (c) month."</p>	

	<p>2011 through the present (the “Relevant Period”) [AGREED]</p> <p>2. Google will perform sampling from 8 days per fiscal quarter during the Relevant Period, with Plaintiffs to select the days. [DISPUTED]</p> <p>3. For each sample day, Google must provide the total number of Records for that 24-hour day, broken down by (a) Canonical Database and (b) Relevant Property. Google will also provide the total number of Records for the entire Relevant Period, broken down by (a) Canonical Database, (b) Relevant Property and (c) month. Plaintiffs’ position is that this information is essential for Plaintiffs and the Court to identify the extent to which the Class Event Data sample reflects the entirety of the Class Event Data. [PARTIALLY AGREED]</p>	<p>Google agrees to do so to the extent possible.</p>	
--	---	---	--

	<p>4. Google will provide a random sample of the Records from both Canonical Databases using the following method: For each sample day, and for each Relevant Property, Google will provide a random sample of 0.5% of the Records for that day.</p> <p>For avoidance of doubt, this means that if there were a combined 2,000,000 Records for one Relevant Property (Property #1) on a sample day, and 1,000,000 Records for another Relevant Property (Property #2) on the same day, Google would produce a random sample of 10,000 Records (0.5%) for Property #1, and 5,000 Records (0.5%) for Property #2 for that sample day. [DISPUTED]</p> <p>5. The parties will meet and confer regarding the method or tool used to select the random sample of Records per sample day.</p>		
--	---	--	--

<p>4. <i>Dates for Completion of Production of Sample Template and Production of all Sample Data</i> [PARTIALLY AGREED]</p>	<p>Prior to making is production of the Class Event Data sample, Google has agreed to provide separate sample templates of the sample data for both Canonical Databases, containing all available fields in the Canonical Databases, in order to allow the parties to resolve any issues with the format of the sample prior to the production commencing. Both sample templates will include 100 Records, with the parties meeting and conferring regarding the method of selection agreed to by the parties.</p> <p>Google has stated that it is able to provide the sample template for the Google Analytics 4 Canonical Database by the end of 2025, but has not yet provided a date by which it will produce the sample template for the Universal Analytics Canonical Database. As Plaintiffs' receipt and review of both</p>	<p>Assuming the Court adopts Google's proposed sample size, Google anticipates being able to complete its production in three months.</p> <p>The systems housing data for Google Analytics are custom-designed for the sole purpose of providing its services and are not set up to externally export the data. Google must secure engineers for the task of building an export solution for this litigation. Additionally, Google's current infrastructure supports limited transfer speed that makes faster production impractical.</p> <p>Google agrees to provide a sample template from the GA4 database by the end of the year.</p>	
---	---	---	--

	<p>sample templates is a precondition to the production of the sample data, Plaintiffs ask the Court to order that Google provide both sample templates by January 5, 2026.</p> <p>Plaintiffs otherwise propose that Google complete its production of the Class Event Data samples within 60 days of the proposed deadline for providing the sample templates, March 6, 2026.</p>		
<p><i>5. Format of Production of Class Event Data</i> [TENTATIVELY AGREED]</p>	<p>Plaintiffs requested that Google produce the Class Event Data samples in a database-ready format, specifically in CSV (UTF-8) format with one row per event/hit and consistent, typed columns. <i>See</i> ECF No. 103 at 14 (ESI Protocol provides “[t]o the extent practicable, the Parties will make reasonable efforts to produce Documents collected from databases in</p>	<p>Google’s will make its production in JSON format. Google reserves its objections to any later request for a production in CSV format.</p>	

	<p>a reasonably useable format. If the data cannot be produced in a reasonably useable format, the parties will meet and confer to address the production format.”).</p> <p>Google has responded that, due to the atypical nature and structure of the Canonical Database, exporting the Class Event Data to CSV format is impractical because doing so would “create a massively wide and massively long CSV that will make the data less usable, not more.”</p> <p>Based on Google’s representations, Plaintiffs are willing to tentatively agree to the Class Event Data sample being produced in JSON format, as Google has requested, but Plaintiffs do not waive their right to seek production in a different format if the JSON files do</p>		
--	--	--	--

	not represent a "reasonably usable format."		
5.a. Fields to be included in Class Event Data [PARTIALLY AGREED]	<p>The parties agree Google's production of the Class Event Data samples will include all fields included in the Canonical Databases, subject to any possible anonymization.</p> <p>Plaintiffs further require that Google include the following fields in its production of Class Event Data, to the extent such information is in Google's possession:</p> <p>(a) Google Account ID field, Google User ID field and the 'joining keys' linking the two¹</p> <p>For data from Universal Analytics Canonical Database</p>	<p>Google disagrees that it should be required to enrich the data by adding additional identifiers and so-called "joining keys."</p> <p>First, the data already includes identifiers (including most but not all of the identifiers Plaintiffs request) that allow plaintiffs to associate records together.</p> <p>Second, enriching the data with "joining keys" to link Google Account ID (i.e., signed-in users) with data from signed-out users is inappropriate and unnecessary, to the extent it is even possible.</p> <p>To start, unlike the pseudonymous identifiers</p>	

¹ See *In re Google RTB Consumer Privacy Litigation*, Case No. 21-cv-02155-YGR (VKD), ECF No. 487 at 7 ("Google must take care that all of its productions of named plaintiffs data can be easily correlated, and it must include the fields necessary to accomplish that result. During the hearing, Google acknowledged that it could include the Google Account ID, Google User ID, and 'joining keys' without undue burden, and the Court expects that it will do so, including remedying its prior productions that omit this information.").

	<p>(a) cid (Client ID)</p> <p>(b) uid (User-ID)</p> <p>(c) session identifiers (e.g., visitId/session number)</p> <p>(d) server/event timestamp,</p> <p>(e) webPropertyId (format UA-xxxxxx-y)</p> <p>(f) fields reflecting mapping between profileId(View ID) and property.</p> <p>For data from Google Analytics 4 Canonical Database</p> <p>(a) user_pseudo_id (derived from _ga)</p> <p>(b) user_id</p>	<p>that exist in the data, Google Account IDs are associated with identifiable individuals. Adding this data thwarts the required anonymization.</p> <p>Further, Google does not join or link signed-in and signed-out data, nor are “joining keys” associated or stored with the data to be produced. Essentially, Plaintiffs ask Google to create evidence that does not exist in the ordinary course.²</p> <p>Finally, Google has already searched for any data associated with the Google Account ID of each named plaintiff. The only data Google could locate was associated with Teresa Wright, who has since dismissed her claims</p>	
--	--	--	--

² Plaintiffs’ reliance on *In re Google RTB* is misplaced. To start, that case involved different data and data sources, so it does not follow that Google can accomplish the same feat here. Additionally, the stated need for identifiers in that case--- to “correlate” named plaintiffs’ data is not present here, as this dispute concerns data regarding *absent* class members. (See Joint Statement, Dkt. 123, at 2 (Plaintiffs distinguishing “class event data” from data concerning named Plaintiffs)). Google has already produced all data associated with named plaintiffs’ identifiers that it has been able to locate.

	<p>(c) ga_session_id (and/or session number)</p> <p>(d) event timestamp</p> <p>(e) numeric property_id</p> <p>(f) measurement_id (format G-xxxxxxx)</p> <p>(g) stream_id</p>	<p>against Google with prejudice. The additional identifiers would not pertain to any named Plaintiff, and it would be meaningless as it relates to class members whose Google Account IDs cannot be confirmed by Plaintiffs.</p>	
<p>5.b. Google's production of records explaining Class Event Data format [PARTIALLY AGREED]</p>	<p>Plaintiffs require Google to produce any documentation it maintains that describes the structure of the Canonical Databases and any other databases from which the Class Event Data samples will be drawn. This includes any available schemas, field definitions, or enumeration tables that identify or explain the data fields contained in those databases.</p>	<p>Google plans to include with its production an explanation of relevant fields within the dataset, which it believes will meet Plaintiffs' needs.</p> <p>Plaintiffs request for "any available schemas, field definitions, or enumeration tables" remains vague after several conferences and is overbroad.</p> <p>Google anticipates that the advance sample it will produce will help the parties narrow down the relevant fields.</p>	

5.c. Agreement on Method of Anonymization of Class Event Data [AGREED]	<p>The parties agree to meet and confer on the method Google will use to anonymize any data in the Class Event Data samples.</p> <p>Google agrees that it will identify and describe any fields of currently non-anonymized data that it intends to anonymize prior to commencing its production of the Class Event Data sample.</p>	[AGREED]	
6. Additional Matters (Data Storage) [DISPUTED]	<p>Plaintiffs have agreed to, have followed, and will continue to follow the requirements of the Protective Order entered in this case (ECF No. 95), including paragraph 12.8 which concerns appropriate safeguards that must be implemented in handling Protected Material like the Class Event Data samples. In light of Plaintiffs' obligations under the Protective Order, Plaintiffs disagree that they are required to make a separate "proposal," ostensibly to</p>	Plaintiffs must provide a proposal for secure and sufficient storage of the requested data.	

	Google (and ostensibly requiring Google's approval), in order to acquire and store this data, in particular where the data will be anonymized per the Court's order.		
--	--	--	--